Estimating the Effects of Time-varying Treatments on Cancer Risk in Randomized and Nonrandomized Studies

Methodological Considerations in Evaluation of Cancer as an Adverse Outcome Associated With Use of Non-Oncological Drugs and Biological Products in the Postapproval Setting

FDA/NCI, September 10-11, 2014

Miguel A. Hernán
Harvard School of Public Health

Disclosure

- The work presented here is supported by NIH grants R01 HL080644 and P01 CA134294

What have been talking about

- Pharmaepidemiological Data and Study Design
- Administrative Databases
- Cancer Registries
- Electronic Medical Record Database
- Biologically relevant time windows
- Exposure determination

What have been talking about

- Pharmacoepidemiological Data Options
- Use of Administ
- Cancer
- [illegible] Design
- [illegible] Record Database
- [illegible] determination, biologically relevant time windows

Observational studies and their problems

What about randomized clinical trials?

- Suppose we have time, money, and IRB approval to conduct very large RCTs
  - to evaluate the effects of non-cancer drugs on cancer risk
- Would that solve all of our problems?
  - The answer is NO

Key problems are shared by randomized/observational studies

- Discussing them in the context of observational studies ONLY may be misleading
- It is helpful to distinguish shared problems from those unique to observational studies
- Let us talk about randomized clinical trials (RCTs) first

Typical pre-approval randomized trials

- Highly-controlled experiments
- Stringently selected Participants
- Short duration
- Small sample size
- No long-term clinical outcomes

- Little deviation from study protocol, high adherence, no losses to follow-up

Typical post-approval randomized trials

- Loosely controlled experiments
- Typical patients
- Long duration
- Large sample size
- Long-term clinical outcomes

- Greater deviations from protocol, low adherence, losses to follow-up
  - Naturalistic, pragmatic, or large simple trials

Very different types of trials

- Pre-approval trials resemble laboratory experiments
- Post-approval trials resemble observational studies
  - Except for baseline randomization of interventions and, perhaps, blinding
  - Benefits of baseline randomization potentially overshadowed by post-baseline noncompliance/loss follow-up?

Research problems in RCTs can be classified into two groups

Those related to

1. Articulating the causal question

and

2. Providing an answer

- Seems kind of silly but it is actually important
  - Let us review the causal questions that can be asked in randomized trials

Causal questions in randomized trials

1. The effect of being assigned to an intervention, regardless of intervention received
   - **Intention to treat** (ITT) effect
   - Interventions to be compared:
     - be assigned to treatment A at baseline and remain in the study until it ends
     - be assigned to treatment B at baseline and remain in the study until it ends
   - Requires adjustment for post-randomization (time-varying) selection bias due to loss to follow-up (Little et al, NEJM 2012)

Causal questions in randomized trials

2. The effect of receiving the interventions specified in the study protocol
   - **Per protocol** effect
   - Example of interventions to be compared:
     - receive treatment A continuously between baseline and study end (unless toxicity arises)
     - receive treatment B continuously between baseline and study end (unless toxicity arises)
   - Requires adjustment for post-randomization (time-varying) confounding/selection bias

Causal questions in randomized trials

3. The effect of receiving interventions other than the ones specified in the study protocol
   - Example of interventions to be compared:
     - receive A as per protocol
     - receive B but switch from B to A if LDL-cholesterol raises above 160 mg/dL (4.1 mmol/L)
   - Requires adjustment for post-randomization (time-varying) confounding/selection bias

Effects vs. *analyses*
The elephant in the room

- Typical ITT and per protocol **analyses**
  - do not adjust for pre- and post-randomization variables
  - Potentially biased estimates of ITT and per protocol **effects**
- Adjustment for post-randomization (time-varying) variables require special techniques
  - Inverse probability weighting, g-formula, etc
    - Developed by Robins et al since 1986
  - Instrumental variable estimation

Intention-to-treat or per protocol effects for post-approval trials?

- Clearly, ITT effect cannot be the default for post-approval, safety trials
  - As recognized by the FDA
- ITT is “conservative” in placebo-controlled trials
  - Unethical when we are trying to estimate effects on cancer
- Per-protocol effect more relevant

But no generally accepted method to estimate per-protocol effects!!

- Typical per-protocol **analysis** is a naïve analysis
  - does not adjust for pre- and post-randomization variables
- We would never accept an observational analysis that does not adjust for pre- and post-baseline confounders
  - Why do we lower standards for randomized trials?

Example of per-protocol effect estimation

- Randomized experiment analyzed like an observational study
- Effect of estrogen plus progestin hormone therapy on risk of breast cancer in postmenopausal women
- Data: Women's Health Initiative randomized clinical trial
  - ~16,000 postmenopausal U.S. women
  - Toh et al. *Epidemiology* 2010; 21:528-539

Methodological challenges for per-protocol effect

- Time-varying treatment
  - Women may not adhere to their assigned treatment (hormone therapy or placebo)
- Time-varying confounders
  - Use of hormone therapy depends on age, BMI, symptoms…
  - may be affected by prior treatment
- Also better to estimate absolute risks
  - Appropriately adjusted survival curves
  - Not only hazard ratios

Estimation of per-protocol effect

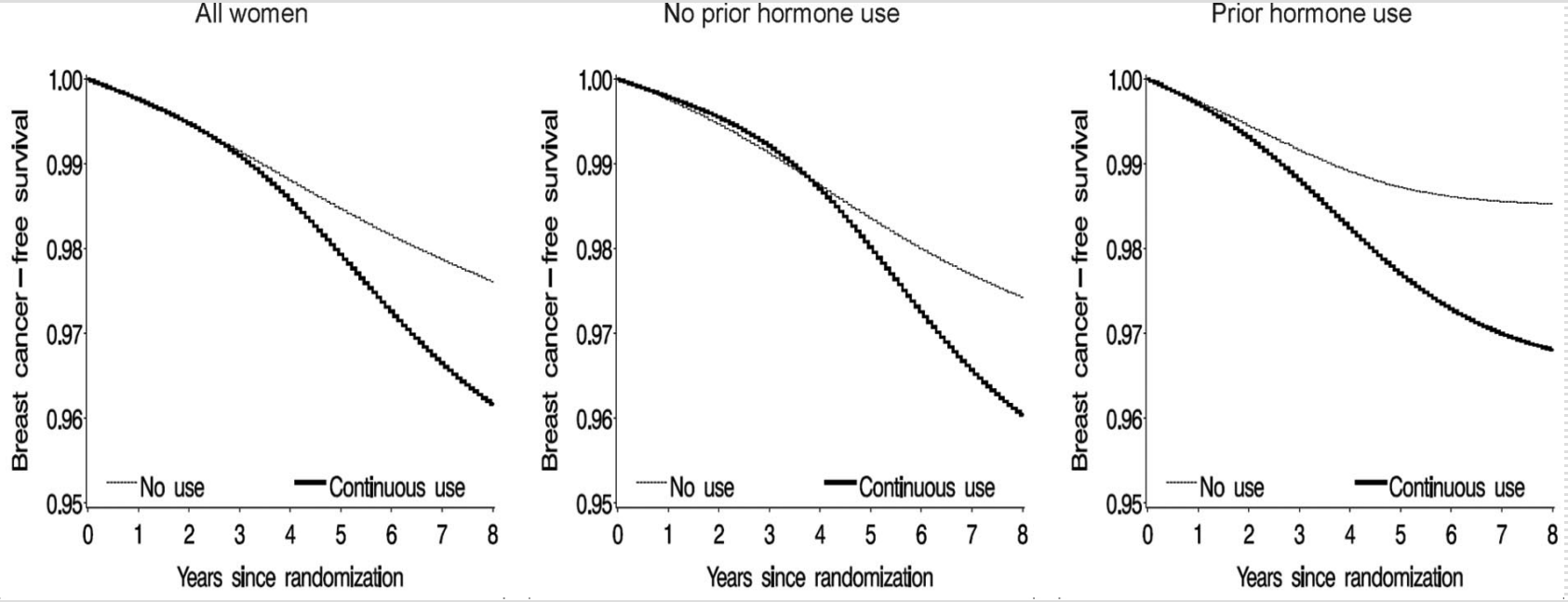
- Estimate stabilized inverse probability (IP) weights to adjust for time-varying confounding
  - Need data on post-randomization variables
- Estimate IP weighted hazards model to estimate
  - Hazard ratios
  - Survival (or cumulative incidence)
- Compare survival curves for continuous treatment vs. no treatment
  - Standardize curves to baseline variables

Hazard ratio of breast cancer
Hormone therapy vs. placebo

- Intention to treat effect
  - 1.25 (1.01, 1.54)
- Per protocol effect
  - 1.68 (1.24 to 2.28)

- Suppose you are a woman considering initiation of hormone therapy and who plans to take it as instructed by your doctor
  - Which hazard ratio do you want?

% free of breast cancer under full adherence to assigned treatment



Toh et al. *Epidemiology* 2010; 21:528-539 (w/ SAS programs)

Research problems in RCTs can be classified into two groups

Those related to

1. Articulating the causal question

and

2. Providing an answer

- We cannot discuss the methods and data necessary to answer #2 until we agree on #1
  - Intention-to-treat or per-protocol?

All of the above applies to observational studies

- Observational studies need adjustment for baseline confounders
  - RCTs do not, at least when they are large
- But, other than adjustment for baseline confounding, **analysis should be identical**
  - both designs need adjustment for time-varying confounding and selection bias
  - because decisions after baseline are not randomly assigned under either design

Post-approval observational studies are our attempt to emulate trials…

… that we cannot actually conduct

- How can discuss the analysis of observational studies with time-varying treatments if we have not agreed on how to analyze the corresponding trials?
  - Observational analyses that adjust for time-varying confounders are exactly equivalent to those of trials that adjust for noncompliance

References

- Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. *International Journal of Biostatistics* 2008; 4(1): Article 22.
- Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease . *Epidemiology* 2008; 19(6):766-779
- Hernán MA, Robins JM. Observational studies analyzed like randomized experiments: Best of both worlds. *Epidemiology* 2008; 19(6):789-792
- Toh S, Hernández-Díaz S, Logan R, Robins JM, Hernán MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology* 2010; 21(4):528-39
- Hernán MA. With great data comes great responsibility. Publishing comparative effectiveness research in EPIDEMIOLOGY. *Epidemiology* 2011; 22(3):290-291
- Danaei G, García-Rodríguez L, Cantero OF, Logan RW, Hernán MA. Observational data for comparative effectiveness research: an emulation of a randomized trial of statins for primary prevention of coronary heart disease. *Statistical Methods in Medical Research* 2013; 22(1): 70-96
- Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: Lessons for comparative effectiveness from a meta-analysis of statins. *American Journal of Epidemiology* 2012; 175(4): 250-262
- Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. *Clinical Trials* 2012; 9(1):48-55
- Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs.* Washington, DC: The National Academies Press, 2012